Prediction of Churn in Telecom Service: Exploring Call Behaviors and using Machine Learning

Jae Won Choi

Abstract*: Churn has a significant impact on mobile network operators and telecommunications service providers. Many studies on churn have been reported, but no one can say that they can create universal human tools for predicting churn or that we can see all the reasons for it. The purpose of this study is to derive the call behavior factors of churning customers and to find ways to reduce the churn of target customers who exhibit these call behaviors. For this, this study uses decision tree and machine learning for the prediction of churn in telecom service. Based on the analysis results, first, the fact that the total number of customers who have more than 316.7 in churn shows that the higher the number of calls, the higher the chance of churn. Second, among customers with total day minutes above 316.7, those with customer service calls above 8.5 show a high likelihood of churn among complaining customers. The overall accuracy is 91.4%. Among the customers who predicted not to be churned, the accuracy that would not be churned was 92.87%, and the accuracy that was churned was 78.4% among the customers predicted to be churned.

***Keywords* : Telecom service; Churn; Decision tree, Machine learning.**

I. INTRODUCTION

Using the correct tool to leader is the correct decisions. As things become more complex and complex, followers are under greater pressure as they anxiously wait for their actions. However, when a leader does not have the right information and data, decision making is often arbitrary and unfounded. The purpose of this study is to provide the communication industry leaders and visionaries with the right information to address one of the industry's problems known as deviations. The procedures used in this study can also be used as sample templates for leaders in all industries to follow on similar issues in customer retention.

Customers who do not renew their contract or look for other service options are referred to as those who leave one airline to another. Churn has caused significant economic and financial losses and is still an important issue for Global Mobile Communications System (GSM) operators despite their attempts to resolve the issue.

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
Jae Won Choi*, College of software, Chungang University, Heugseoglo 84, Dongjaggu, Seoul, Korea, e-mail: ch000523@naver.com

Prediction of software is developed in order to solve the problem. The result of the software is quite correct. Nevertheless, the software uses statistical equations to predict future departures without clear indication of small causes.

Because change is a fact of life, deviations are a problem not only for the telecommunications industry, but for all industries facing retail, service and competition and market saturation [1]. To suddenly change and do not know the cause of change management is inevitable, but the problem is a terrible mistake. Determine the cause of this quantitative study of the telecommunications industry and should support. The methodology used in this paper can be regarded as a roadmap for the reader to follow the steps taken in this study and to apply the one-day procedure to identify the causes of many other problems. The purpose of this study is to find and analyze the causes of hope so that communication industry leaders and managers can use the study to formulate possible solutions to problems.

II. RELATED STUDY

A. Churn in the telecommunication industry

Customers have become more intelligent and have access to a variety of options. Price is an important factor for customers to consider. Churn has a significant impact on mobile network operators and telecommunications service providers. TW Telecom reported $2.4 million in lost revenue and a 6.6 percent drop in network services due to churns from contract renewals and price increases [2]. AT&T, a well-known wireless operator, reported a bitter crisis for its rival T-mobile due to its fight against the spectrum. AT & T had to lose its spectrum allocation, and either couldn't deliver it or had to offer limited services to its customers. When a customer tries to find the superior network services offered by a competitor, this results in an immediate churn. AT&T reported a slight increase from 1.32 percent in the fourth quarter of 2010 to 1.36 percent in the third quarter [3]. Even Twitter considers the impact of deviations on services. Within a month, 60% of Twitter accounts were reported to be idle. Twitter reported that it was looking for ways to improve relevance and user experience by adding new services and features.

B. The determinants of churn

A lot of researchers have suggested many kinds of determinants of churn in telecom service as the following table 1.

<Table 1> The determinants of churn

Category	Contents	Reference
Price	Customers are based on usage and price sensitivity using the Customer Segment Model (CSM) and customer-created segment features.	[4]
Competition	Competition between existing companies and new entrants using an economic assessment has been reported as a price war if a new entrant attempts to penetrate the telecommunications market and obtain market share from existing providers.	[5]
Profitability	Keeping an existing customer is much cheaper than having a new one, resulting in financial benefits.	[4]
Loyalty	The potential for loyalty depends on three environments: (1) the product or supplier, (2) the customer principal, and (3) the market and other suppliers.	[5]
Marketing	Deviance is around 25%, a seasonal high of around 25%, and this rapid increase is usually caused by advertising campaigns run by mobile network operators during summer vacations and after-school hours.	[6]
Branding	New innovative services help rural communities. These services also promote the brand and image of the operator as a social responsibility operator.	[7]
Organizational Policies	Exploitation strategies such as geographical expansion and delivery of new services have prevented a breakaway.	[5]
Technological Development	The introduction of 3G technology has led to an increase in churn at a much higher rate than mobile portability, in line with other studies that technology evolution is a key player in breaking away.	[6]
Number Portability	According to a study of Korea's telecommunications market, 40% of customers are willing to leave if they can maintain their current number.	[8]
Network Quality	Customers who find their services unreliable tend to leave.	[1]

Many studies on churn have been reported, but no one can say that they can create universal human tools for predicting churn or that we can see all the reasons for it. Ginn et al. [1] reported that the indicators are customer surveys and customer indicators such as customer satisfaction index, customer expectation index, customer value analysis, brand preference analysis, repurchase and recommendations, and walllet analysis and expenditure trends.

While most of the literature analyzed only age and gender, only a small number of marriages and geographic locations were examined. There is numerous literature that has tested the impact of various call plans and deviations. However, there is no literature linking these demographics to plans to convene. Similarly, although new and non-communication services have been reported to reduce customer churn, no literature has examined the churn of a particular service individually. Churn is so complex and linked to so many factors that researchers tend to take a small number of factors and ignore the effects of others. There are up to 250 different factors that can affect a breakaway, but even the best software chooses which of the 30 to 50 elements to test. This paper is to understand what factors should be taken into consideration by the leader with the aim of qualitative methods. Wong [9] said the breakaway study was incomplete because it could be affected by a wide range of factors. Because wireless carrier data is not available, some factors, such as income level, training background, and marital status, are not included. Customer demographics are often changed, so that customers' demographics are constantly monitored, causing problems with the mobile network operator, which can compromise their personal information. Some studies looked at age, gender, and geographic location. But researchers still can't express cultural and behavioral factors that can affect churn.

III. METHODOLOGY

A. Research framework

There are two approaches to customer departure management: The first is that carriers rely on superior products and mass advertising to increase brand awareness, loyalty and customer retention. The second is to identify and deliver services that are likely to break away with incentives or offers from mobile network operators who can customize their service plans to meet their needs. There are two approaches to departure identification: The first response, this will be the carrier calling to complain, and once the carrier is identified as a valued customer, it will try to find the customer again. The other is the operator's active approach to identify customers most likely to leave the customer through Customer Departure Prediction software. The purpose of this study is to derive the call behavioral factors for customer churn and to find ways to reduce the deviations of the target customers that represent these call behaviors.

B. Dataset

The variable information is obtained from the popular Internet platform, Kagle (www.kaggle.com) and from the third-party Web site, the international Web site.This website provides data titled 'Churn in Telecom's Data Set' uploaded by David_beaks. Kagle asked participants to predict the change in communication services. To help develop algorithms, the organizers provided data stream types for large call actions.

Based on the relevant literature on churn based telecom service, the effects of 17 variables are examined. These variables are listed and defined in Table 1.

< Table 2> The variables in each category

Categories	Variables
Churn	Churn
Call behaviors	account length, international plan, voice mail plan, number voice mail, messages, total day minutes, total day calls, total day charge, total eve minutes, total eve calls, total eve charge, total night minutes, total night calls, total night charge, total intl minutes, total intl calls, total intl charge, customer service calls

C. Analysis method

This study uses RapidMiner tool to do decision tree and machine learning for the prediction of churn in telecom service.

Decision tree models organize decision rules. In the decision tree, each node represents an attribute used in a decision, and a branch represents a condition applied to a decision. The leaf node at the end of the tree represents the decision that concludes. In general, a decision tree consists of two processes: growing and pruning. In the growing phase, a decision tree is obtained by storing the appropriate split criteria and stop rules according to the purpose and data structure of the analysis. In the pruning phase, on the other hand, it is the step of eliminating branches that have a high risk of misclassification or have inadequate rules. The most important part of decision tree growth is the selection of attributes for decision making. Known optimal properties include information gain, information gain ratio, Gini index, and classification error.

In order to survive the competitive market, many companies use data mining technology to analyze decision-making predictions. To effectively manage your customers, it's important to build a more effective and accurate model for forecasting decisions. Statistical and data mining techniques were used to construct decision prediction models. Data mining techniques can be used to detect interesting patterns or relationships in data and to predict or classify behavior by fitting models based on available data. In the case where the learning dataset and the test dataset are separated for machine learning, the test dataset must satisfy the following requirements. First, the training dataset and the test dataset must be created in the same format. Second, the test dataset should not be included in the training dataset. Third, the training dataset and the test dataset must be consistent in data. However, it is very difficult to create a test data set that meets these requirements. In data mining, various verification frameworks using one dataset have been developed to solve this problem. This study uses the Split Validation operator provided by RapidMiner to support this. The operator splits the input dataset into a training dataset and a test dataset to support performance evaluation. This study selects relative segmentation among the segmentation method parameters of this operator and uses 70% of input data as learning data.

Performance assessment uses training data to determine how well the generated model works. Performance measures can be divided into technical performance measures and heuristic measures. The technical performance measures to be used in this study show performance results by generating models from training data, processing test data into models, and comparing the class labels of original verification cases with predicted class labels. Measuring technical performance can be divided into supervised and unsupervised learning. The supervised learning used in this study is classified and regressed. The data used for this learning and test all have original class values. The performance is obtained by comparing and analyzing the original class values with the prediction results.

The classification problem is the most common data analysis problem. Various metrics have been developed to measure the performance of classification models. For classification problems of category type, accuracy, precision, recall, f-measure are used a lot. RapidMiner includes Performance (Classification), which measures performance indicators for common classification problems, and Performance (Binominal Classification), which provides performance indicators specific to binomial classification problems. The table below shows how these indicators are calculated.

<Table 3> Key performance indicators

		Actual class (as determined by Gold Standard)	
		True	False
Predicted class	Positive	True Positive	False Positive (Type Ⅰ error)
	Negative	False Negative (Type Ⅱ error)	True Negative

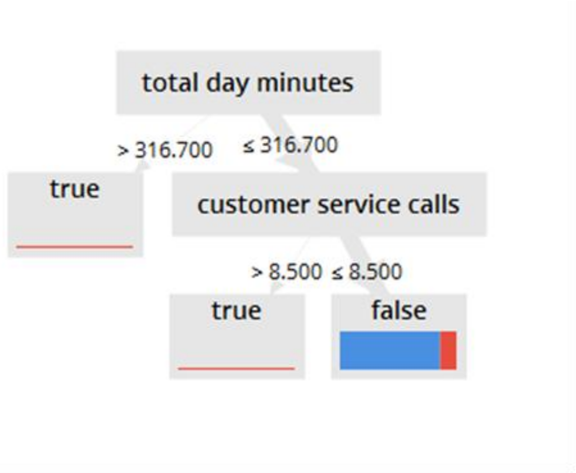
Precision = TP/(TP+FP), Recall = TP/(TP+FN), True negative rate = TN/(TN+FP), Accuracy = (TP+TN)/(TP+TN+FP+FN), F-measure = 2・ ((precision・recall)/(precision + recall))

IV. RESULTS

A. Decision tree

The decision tree analysis results are as follows.

<Figure 1> The results of the decision tree



The analysis shows that customers with total day minutes of more than 316.7 are churning. Also, among customers whose total day minutes are not more than 316.7, those with more than 8.5 customer service calls are churning. These results show that the customer's call behavior can be used to predict churn. Based on the analysis results, first, the fact that the total number of customers who have more than 316.7 in churn shows that the higher the number of calls, the higher the chance of churn. Because call-intensive customers will be concerned about phone charges, they will consider churn to other carriers that offer lower rate plans or discount rates based on call volume.

Second, among customers with total day minutes above 316.7, those with customer service calls above 8.5 show a high likelihood of churn among complaining customers. Customers who file a lot of complaints will consider churn to other carriers emotionally because they are often inconvenient or sensitive to discomfort.

B. Performance evaluation

The performance evaluation results are as follows.

<Table 4> Performance evaluation

	True false	True true	Class precision
Pred. false	833	64	92.87%
Pred. true	22	81	78.64%
Class recall	97.43%	55.86%	

The overall accuracy is 91.4%. Among the customers who predicted not to be churned, the accuracy that would not be churned was 92.87%, and the accuracy that was churned was 78.4% among the customers predicted to be churned

V. CONCLUSIONS

This study identifies the factors determining churn based telecom service. Churn has a huge impact on mobile network operators and telecom service providers. A lot of studies have been reported about deviations, but no one can say that we can create a universal human tool to predict deviations, or that we can all know why. Churns are so complex and connected to so many elements that researchers tend to use fewer elements and ignore the effects of other factors. This paper aims to use quantitative methods to understand the factors that leaders should consider. Customer demographics are often changed, and customer demographics are constantly monitored, which can cause problems with mobile network operators and compromise personal information. Some studies looked at age, gender, and geographic location. But researchers are still unable to express cultural and behavioral factors that can affect deviance. The methodology used in this paper can be viewed as a roadmap for the reader to follow the steps taken in this case study and to apply the one-day procedure to identify the causes of many other problems.

This study, along with Kagle's reining, is leading the way in using data sets to explore the deviations determinants of communication services. The results provide a comprehensive understanding of the reasons for the departure decision of the communication service. Based on a limited set of capabilities, including customer telephone behavior, this white paper proposes the highest performance model for predicting a communication service departure. Use machine learning techniques, including decision tree and neural network, along with functional criticality analysis to get the best results in terms of accuracy. Through this methodology, this study identified a pattern of call behavior that could predict a customer's departure. This study contributes to the literature on communication service variation by providing a global model that summarizes the variability in customer call behavior. In fact, this study provides stakeholders such as telecom service providers with insight to manage the possibility of departure and improve revenue. In addition, this study can provide specific work instructions to the carrier practitioners who are trying to prevent deviations by quantifying the decision factors that actually occur.

Nevertheless, this study recognizes important limitations of this study. Economic modeling is used to explore data sets and identify associations between various factors and deviations. However, social or psychological factors that control customer churn cannot be considered. Therefore, it is important to carry out a sexual study in order to find a theoretical basis for deviations. Future studies of this study include (i) the study of other functional selection schemes, such as the importance of random forest function, (ii) additional experiments on neural network architectures, and (iii) the performance of K- average clustering, especially using the Ridge Regression model.

REFERENCES

1. J. Ginn, M. Stone, and Y. Ekinci, "Customer retention management in the recession. Journal of Direct," Data and Digital Marketing Practice.12, 2010, pp.115– 127.
2. J. G. Barr, TW telecom Company Profile. Faulkner Information Services, 2012, Retrieved from. http://www.faulkner.com.ezproxy.apollolibrary.com/products/faccts/
3. R. Drumheller, "AT&T Company Profile. Faulkner Information Services, 2013, Retrieved from. http://www.faulkner.com.ezproxy.apollolibrary.com/products/faccts/
4. Y. Chen, G. Zhang, D. Hu, and C. Fu, "Customer segmentation based on survival character," Journal of Intelligent Manufacturing,18, 2007, pp.513–517.
5. D. Katsianis, A. Gyurke, R. Konkoly, D. Varoutas, and T. Sphicopoulos, "A game theory modeling approach for 3G operators," Netnomics, 8, 2007, pp.71-90.
6. E. Sutherland, "Mobile number portability," Info. 9, 4, 2007, pp.10-24.
7. The economist, Telecoms in Emerging Markets. The economist.com, 2009, Retrieved from http://www.economist.com.ezproxy.apollolibrary.com/specialreports?year[value] [yea r]=2009&category=76984
8. M. Park, D. Kim, and S. Lee, "Demand for Number Portability in the Korean Telecommunications Market," Global Information Management, 15, 1, 2007, pp.43-64.
9. K. K. Wong, "Getting what you paid for: Fighting wireless customer churn with rate plan optimization," Database Marketing & Customer Strategy Management, 18, 2, 2011, pp.73–82.

AUTHORS PROFILE

Jae Won Choi, Divison of software, College of software, Chungang University, Seoul, The republic of Korea. My interests in studying are data science, artificial inteligence, blockchain, game, etc.